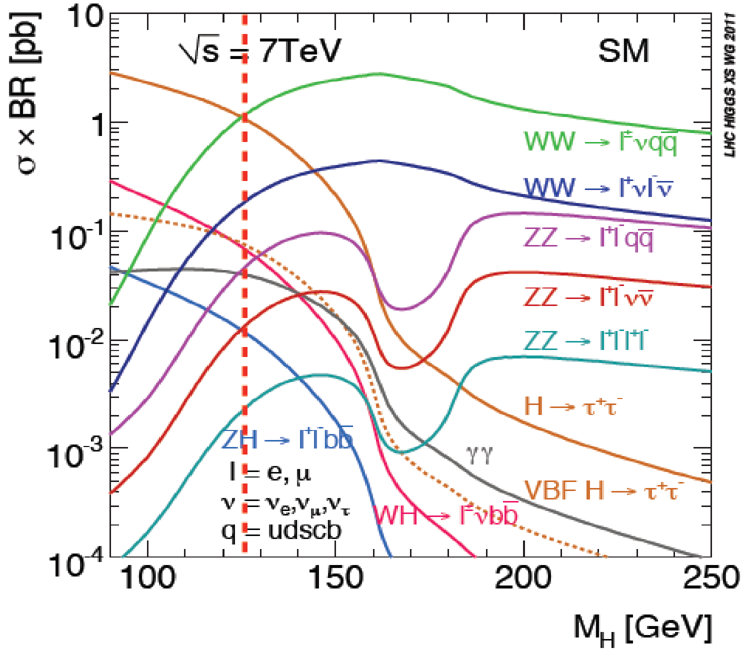
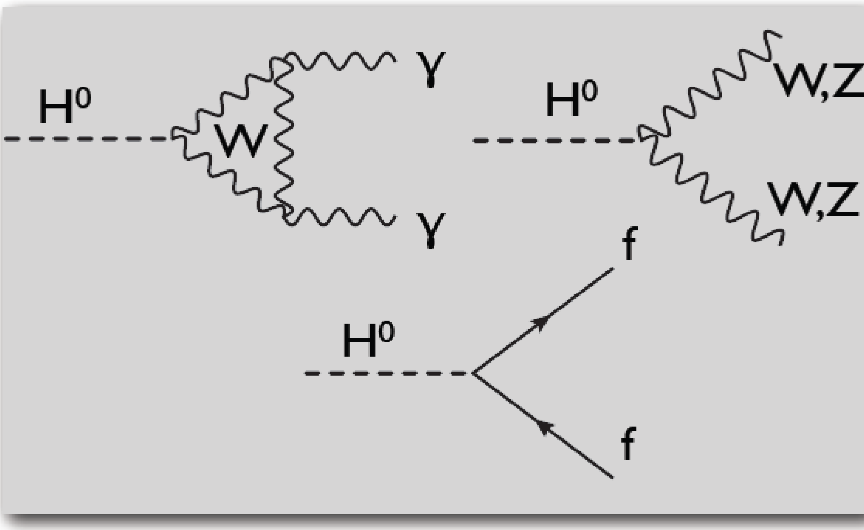
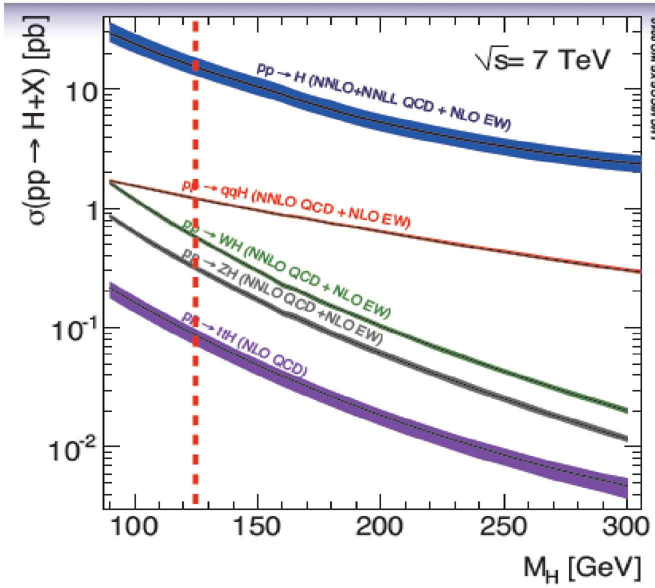
Motivation: $H \to \tau\tau$

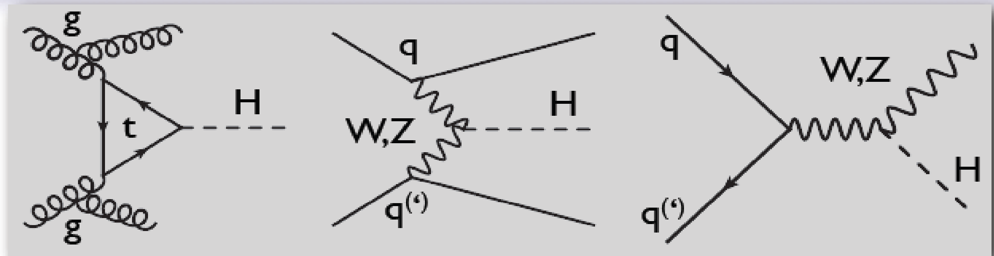




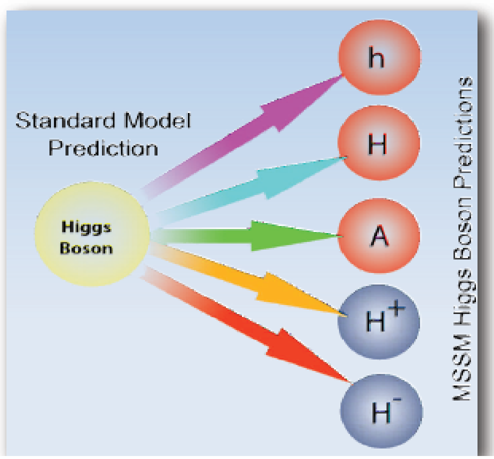
- $H \to \tau\tau$: one of the largest branching fractions for m_H = 126 GeV
- direct access to Yukawa couplings, giving masses to fermions
- in MSSM, enhanced coupling to τ as a down-type fermion ($\tan\beta > 1$)

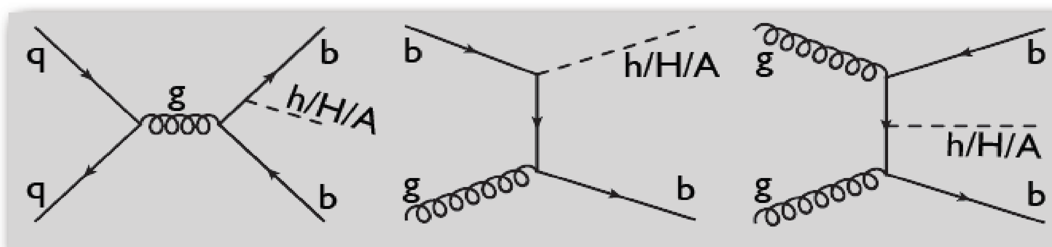
Higgs production in SM & MSSM





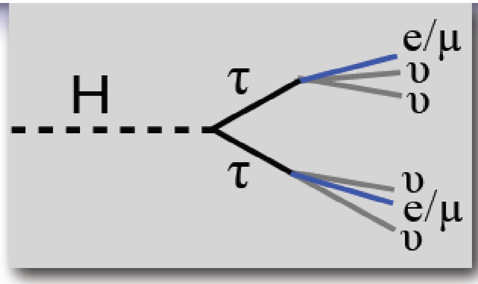
- In SM three main production channels:
  - gluon-fusion (ggH), VBF (qqH), associated (VH)
  - can exploit specific jet topologies in analysis



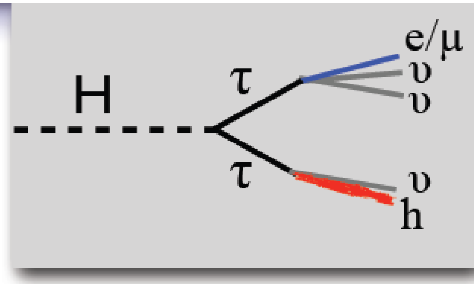


- In MSSM: 2 CP-even (h,H), 1 CP-odd (A) “Higgs”
  - can fix Higgs sector by m_A, **tan β** at Born-level
  - enhanced coupling to b-quarks: exploit b-tagging

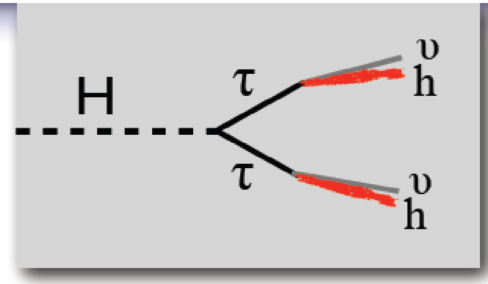
H→ττ Final States



H→ $\tau_{lep}\tau_{lep}$→ ll4ν (12.4%)

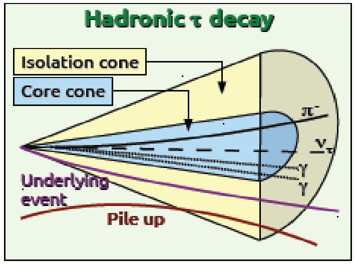


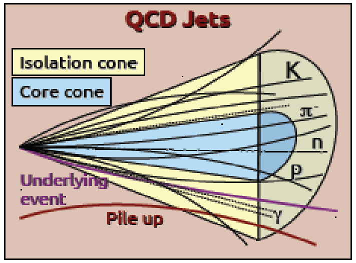
H→ $\tau_{lep}\tau_{had}$→ lh3ν (45.6%)



H→$\tau_{had}\tau_{had}$→hh2ν (42.0%)

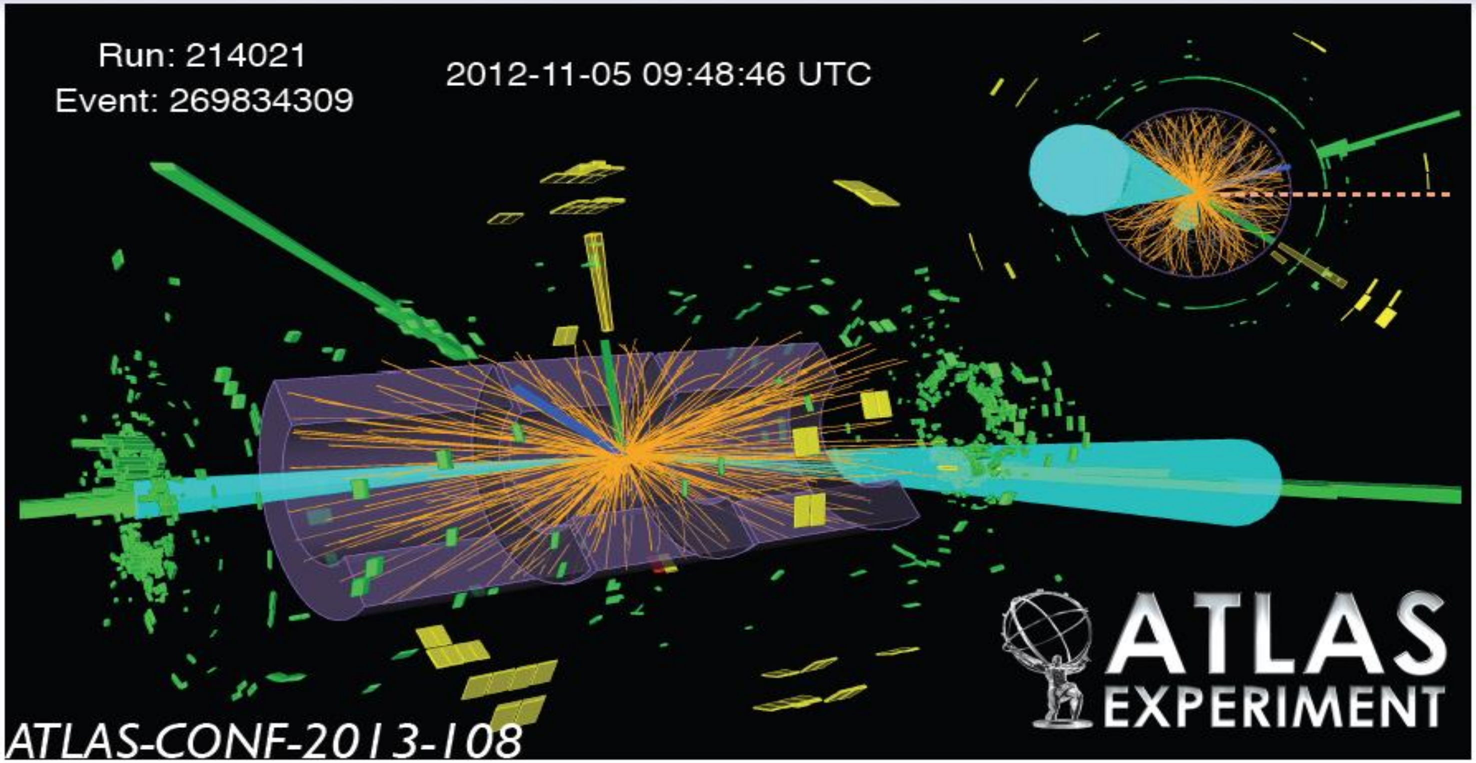
- τ-leptons subsequently decay into e, μ or hadronically: τ_{had}
- problems in pp-collisions





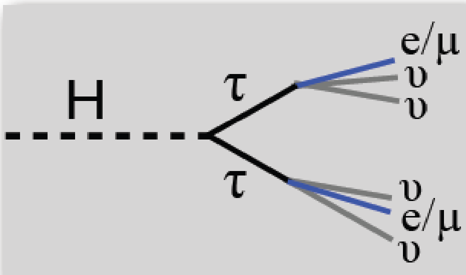
  - need to distinguish "τ-jets" from QCD-jets (see talk by C. Limbach)
  - can only access transverse component of ν momentum sum ($E_{T,miss}$)
- separate analyses for 3 final states (channels): $\tau_{lep}\tau_{lep}$, $\tau_{lep}\tau_{had}$, $\tau_{had}\tau_{had}$
  - different triggers, background compositions, suppression cuts

Candidate Event: (qq)H→ττ→eτ$_{had}$



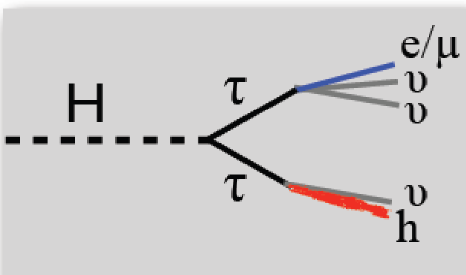
- $p_T(e)$ = 56 GeV, $p_T(\tau_{had})$ = 27 GeV, $E_{T,miss}$= 113 GeV, $m_{\tau\tau}$(MMC) = 129 GeV
- two forward jets with m_{jj} = 1.53 TeV - assume VBF production
- very pure analysis selection with S/B=1.0, 50% chance to be a real signal event

SM Higgs Preselection



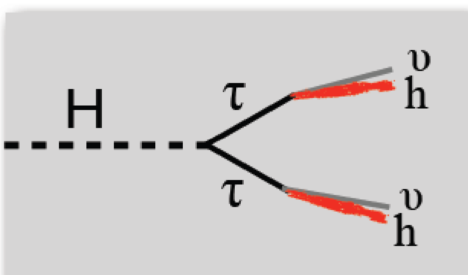
$\tau_{lep}\tau_{lep}$: two isolated leptons with opposite-sign (OS) charges (ee,μμ,eμ)

- ***trigger on leptons*, *single e or μ*** (p_T > 24 GeV) or combined ***ee, μμ, eμ***
- suppress ***Z(ee/μμ)+jets: select*** 30 < ***m(ll)< 75 GeV***, $E_{T,miss}$> 40 GeV
  - leptons ***not back-to-back: Δφ(l,l) < 2.5***, $p_T(l_1+l_2)$ > 25 GeV
  - for eμ: looser cuts 30 < m(ll) < 100 GeV, $E_{T,miss}$> 20 GeV



$\tau_{lep}\tau_{had}$: isolated lepton (e,μ) and „medium" τ_{had} with OS charge

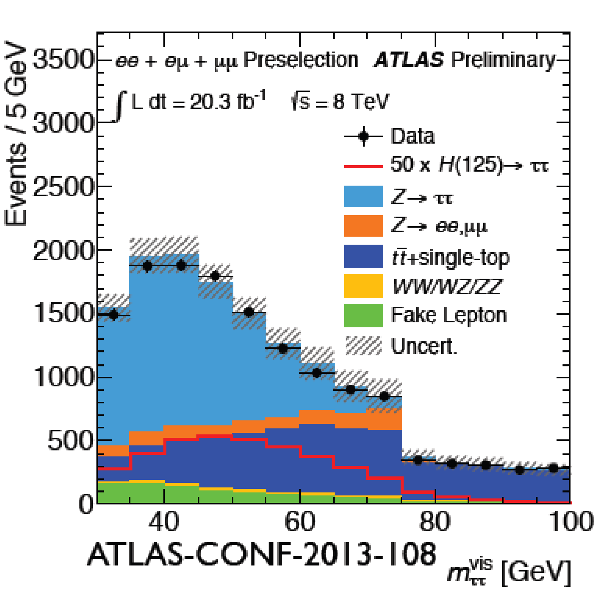
- ***trigger on single e or μ*** (combined e/μ+τ_{had} triggers not included)
- suppress ***W+jets*** by m_T(e/μ, $E_{T,miss}$) < 70 GeV $m_T = \sqrt{2 p_{T1} p_{T2} (1 - \cos\Delta\phi)}$

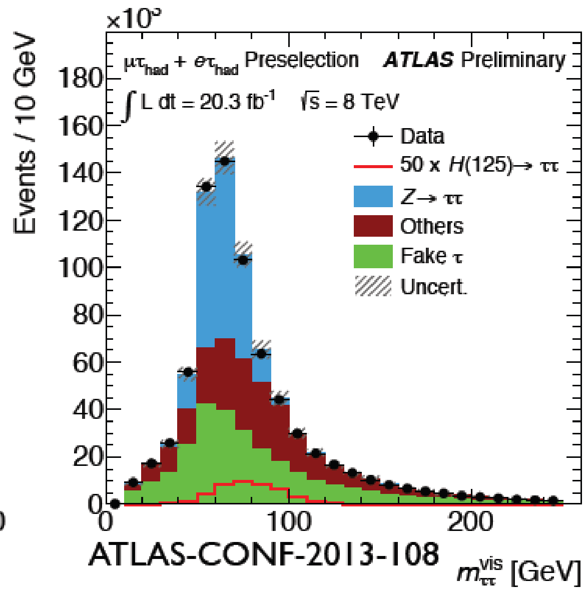


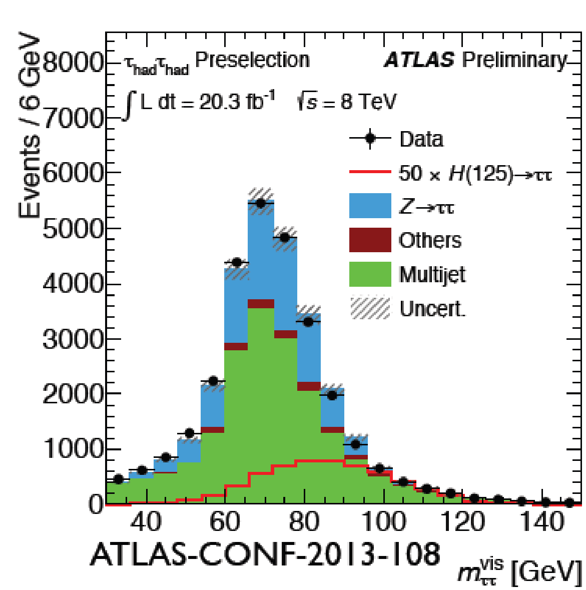
$\tau_{had}\tau_{had}$: one „tight" τ_{had} and one „medium" τ_{had} with OS charge

- ***trigger on di-τ_{had} events*** (p_T > 29 (20) GeV) - ***offline: p_T > 35 (25) GeV***
- suppress ***multi-jet*** events: 0.8 < ΔR(τ_{had},τ_{had}) < 2.8, Δη(τ_{had},τ_{had}) < 1.5
  - require $E_{T,miss}$ > 20 GeV between the τ_{had}: ***min[Δφ(τ_{had},$E_{T,miss}$)]<π/2***

SM: Background Estimation





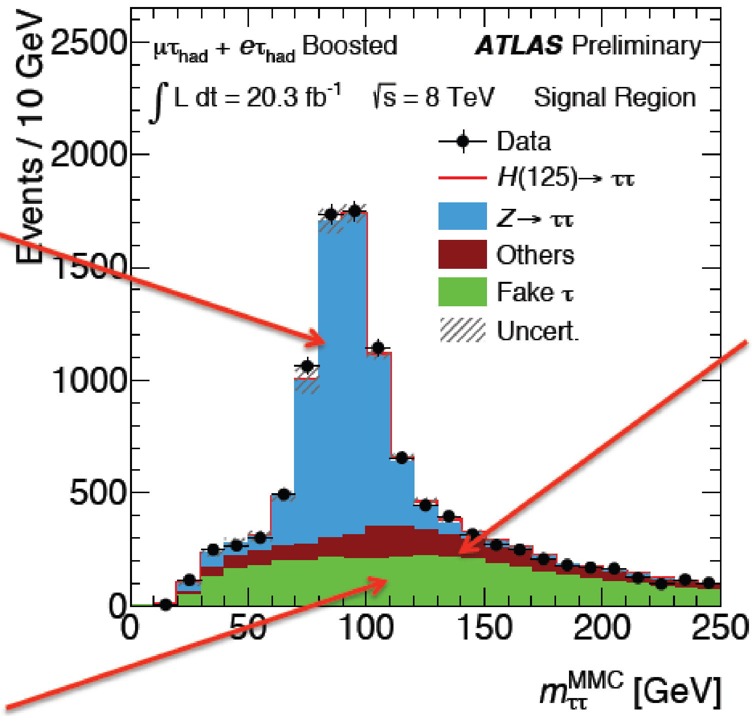


- Z→ττ ***dominant background***, important to be data-driven
  - ***but***: irreducible, no signal-free control region by construction
  - start from ***Z→μμ data***, replace muons by ***simulated τ-decays***: “embedding”

SM: Background Estimation

- All major backgrounds are either directly estimated from data or normalized to data in dedicated control regions

Z→ττ: major background; modeled by data



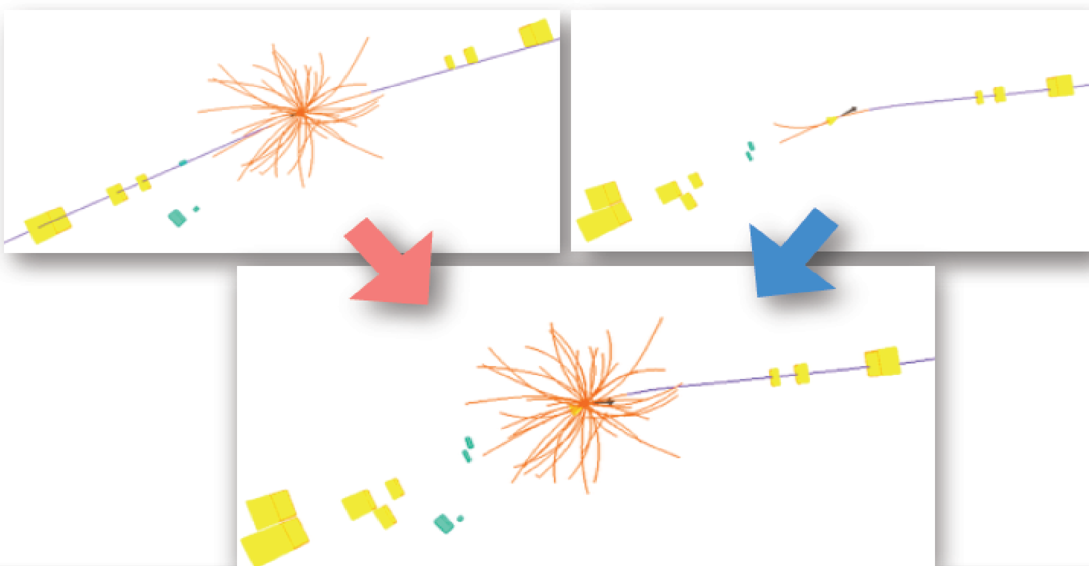
"Others": Dibosons & H→WW* modeled by MC; Z→ee/μμ & top modeled by MC and normalized to data

"Fakes": multijet, W+jets, top (with fake tau); modeled by data

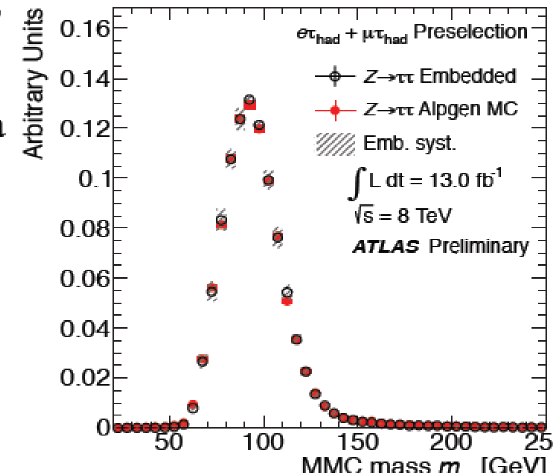
Estimation of $Z\rightarrow\tau\tau$ Background

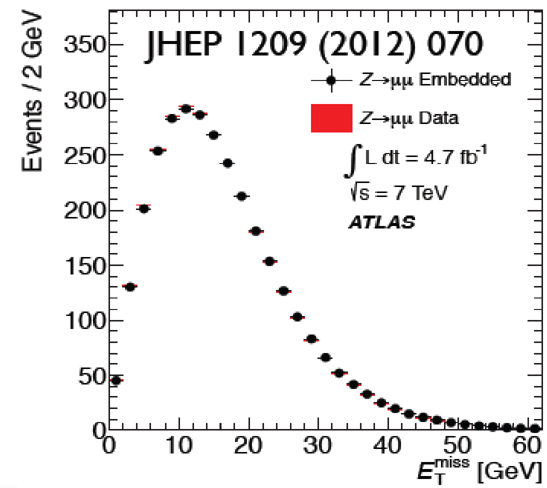
Important shape of $Z\rightarrow\tau\tau$ from $Z\rightarrow\mu\mu$ data: "***Embedding***"

- remove muon tracks and simulated calorimeter energy from data
- replace by full-simulated $Z\rightarrow\tau\tau$ decays, generated with Tauola
- re-run full event reconstruction: pile-up, jets and $E_{T,miss}$ from data
- validation with Monte Carlo and $\mu\rightarrow\mu$ embedding (data to data)

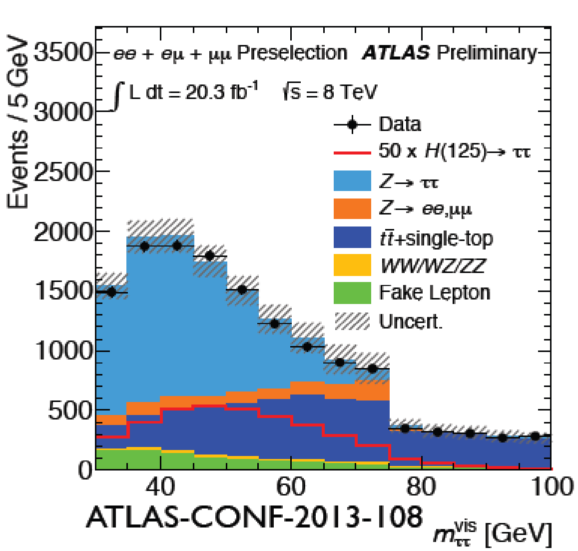


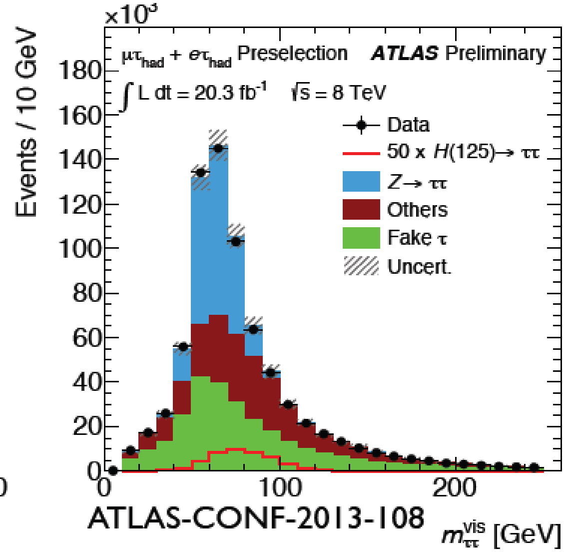
ATLAS-CONF-2012-160

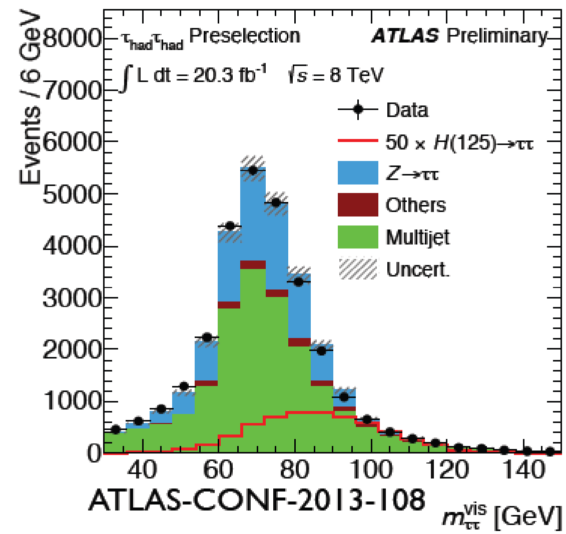




SM: Background Estimation

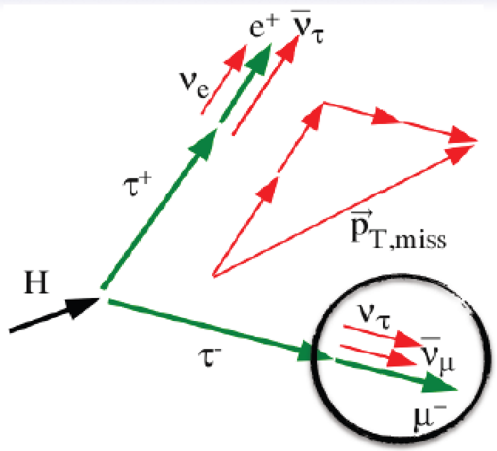


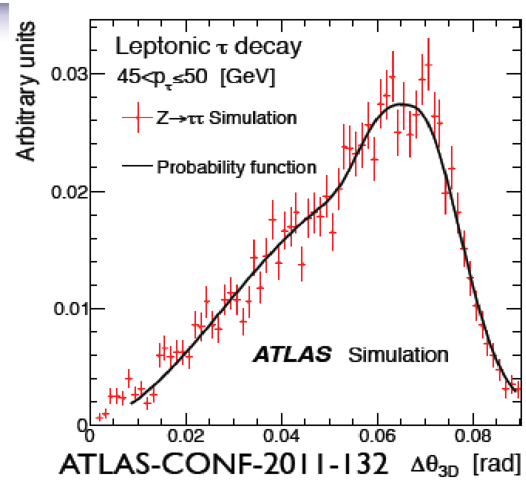




- Z→ee/μμ (+jets), top, di-boson from Monte Carlo simulation
  - corrections to data from dedicated control regions, normalisation from fit to the data
- Multi-jet (and W+jets) from mis-identified QCD jets as lepton or hadronic τ-decays
  - $\tau_{lep}\tau_{lep}$: ***invert isolation, fit multi-jet template*** to data in sub-leading lepton p_T distribution
  - $\tau_{lep}\tau_{had}$: jet→τ_{had} "***Fake-Factors***" from multi-jet and W+jets control regions, ***apply to anti-τ data***
  - $\tau_{had}\tau_{had}$: ***invert OS*** requirement, ***fit multi-jet template*** together with Z→ττ to data in Δη(τ,τ)

Mass reconstruction from E_T^{miss}

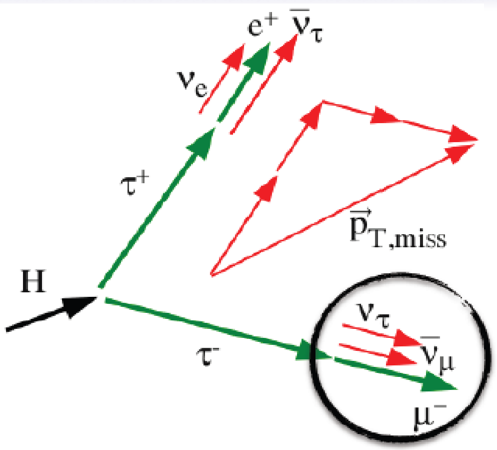


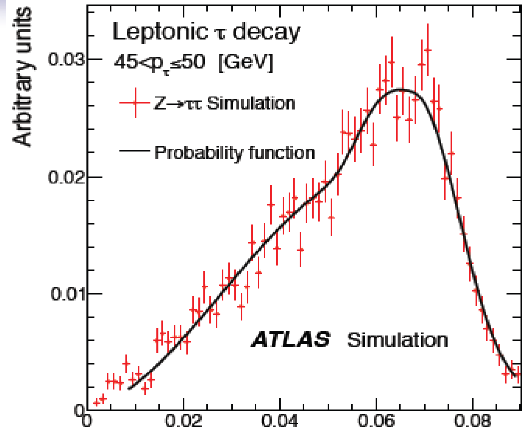


ATLAS-CONF-2011-132

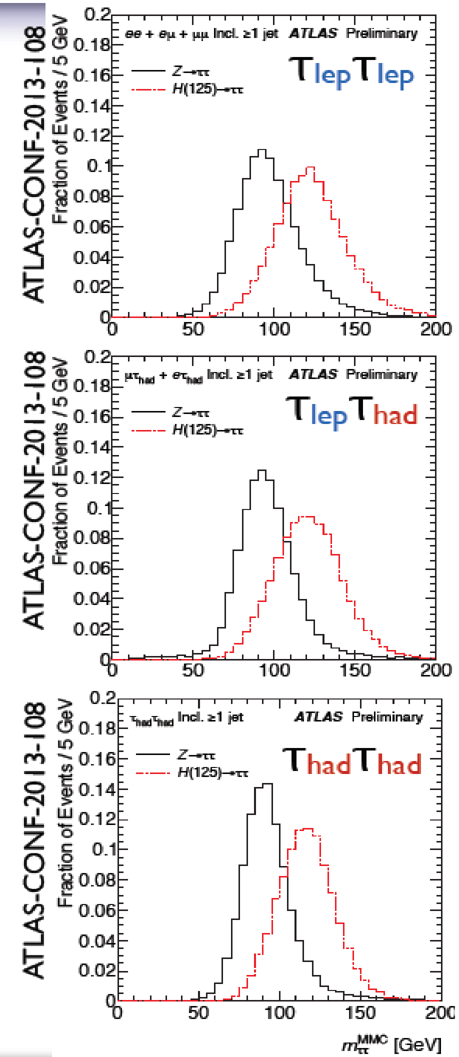
- ***Basic Idea: Collinear Approximation: $\Delta\theta_{3D}(\tau_{vis},\nu)=0$***
  - project $E_{T,miss}$ on visible τ-decay products ($0<\Delta\varphi(\tau,\tau)<\pi$)
  - require visible momentum fractions $0 < x_{1,2} < 1$: $m_{\tau\tau} = \frac{m_{vis}}{\sqrt{x_1 x_2}}$
- ***MMC**: Missing Mass Calculator*
  - $\Delta\theta_{3D}(\tau_{vis},\nu)$ ***from simulation*** ➔ weights in $\Delta\varphi(\tau_{vis},\nu)$
  - most probable value as MMC mass estimator
  - improves resolution, efficiency ($0<x<1$), esp. for ττ back-to-back
  - based on A. Elagin et al., NIM A654 (2011), 481-489, arXiv:1012.4686

Mass reconstruction from E_T^{miss}





ATLAS-CONF-2011-132



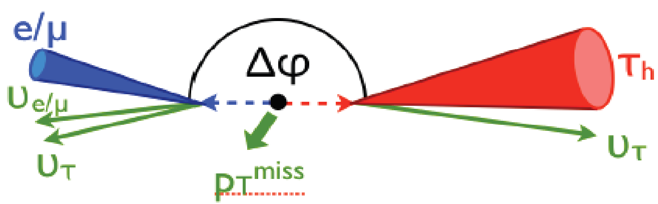
- ***Basic Idea: Collinear Approximation: $\Delta\theta_{3D}(\tau_{vis},\nu)=0$***
  - project $E_{T,miss}$ on visible τ-decay products ($0<\Delta\varphi(\tau,\tau)<\pi$)
  - require visible momentum fractions $0 < x_{1,2} < 1$: $m_{\tau\tau} = \frac{m_{vis}}{\sqrt{x_1 x_2}}$
- ***MMC: Missing Mass Calculator***
  - $\Delta\theta_{3D}(\tau_{vis},\nu)$ *from simulation* → weights in $\Delta\varphi(\tau_{vis},\nu)$
  - most probable value as MMC mass estimator
  - improves resolution, efficiency (0<x<1), esp. for ττ back-to-back
  - based on A. Elagin et al., NIM A654 (2011), 481-489, arXiv:1012.4686

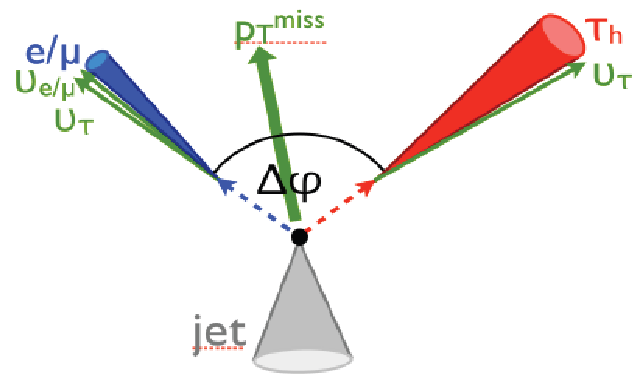
SM: Multi-Variate Analysis Strategy

- *After preselection, define two loose categories per channel:*

- VBF: at least two additional jets (p_T>30-50 GeV) with large separation in $\Delta\eta$ >2.0-3.0
  - exploit qqH production mode, where the two quarks produce „forward“ jets

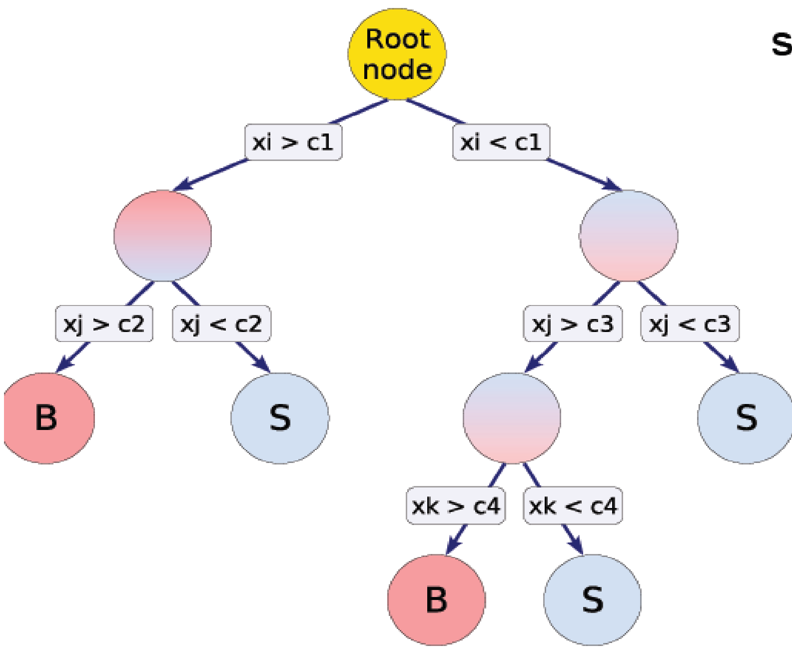
- Boosted: large resonance transverse momentum **$p_T(\tau_{vis}+\tau_{vis}+E_{T,miss})$ > 100 GeV**
  - only consider events which do not fulfil the VBF criteria
  - exploit additional boost of ττ-system, event topology and mass reconstruction





- *For each category and channel, train a* **Boosted Decision Tree (BDT)**

Boosted Decision Trees



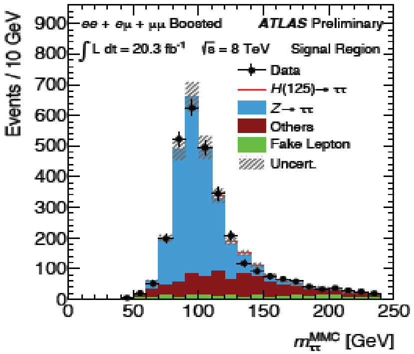
during training, signal and background are known, can extract fraction of mis-classified s in B or b in S

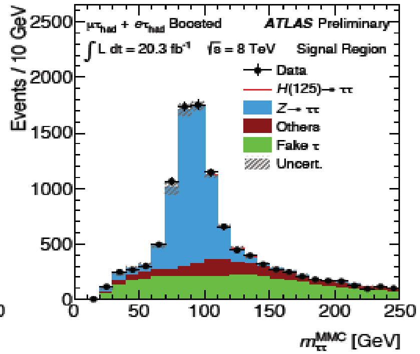
- Multi-variate technique can "learn" how to separate signal from background(s)
  - "Training" builds tree of decisions (cuts) on given variables to separate S from B events
  - subsequently using variables with large separation power to define many hyper-cubes as S- or B-like

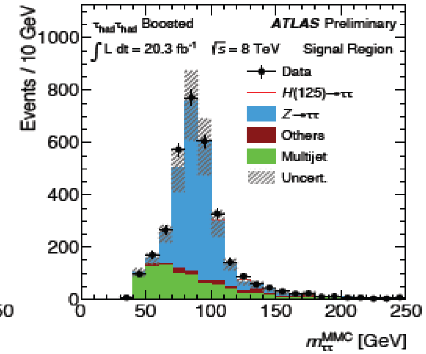
  ⇒ *so far "only" a highly optimized cut selection*

  - use weights for mis-classified events and train new decision tree - learn from mistakes!
  - can train O(1000) trees which are "***boosted"*** by using these ***weights***
  - final output ("BDT score") for each event: purity-weighted sum over all trees projected on [-1,+1] (-1 for background, +1 for signal)

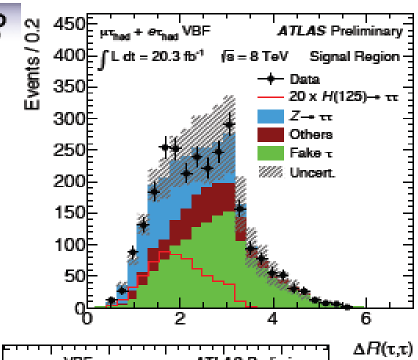
SM: Example Observables for BDTs

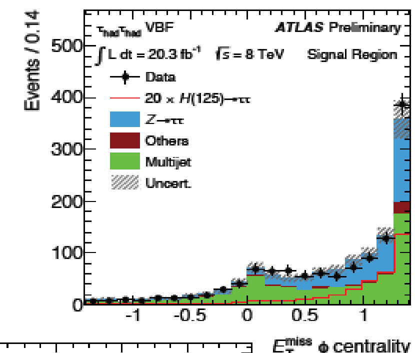


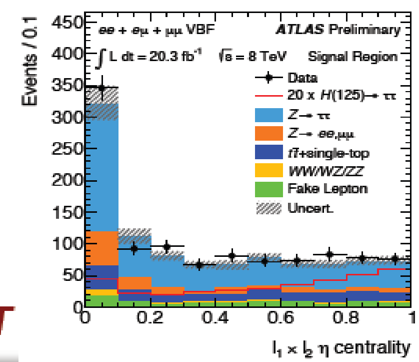




ATLAS-CONF-2013-108



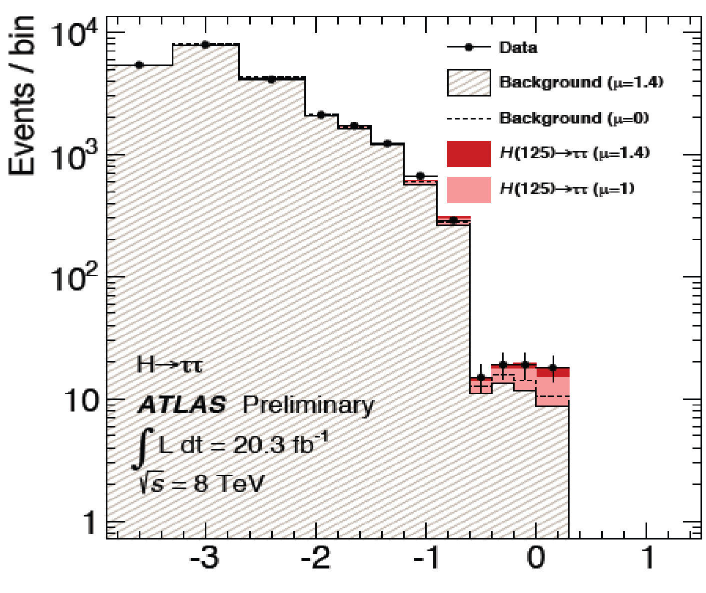




- Selection of BDT input variables harmonised, still certain differences between channels (and categories)
  - *MMC mass* and $\Delta R(\tau_{vis},\tau_{vis})$
    - also $\Sigma_{(\tau,\tau,j)} |p_T|$ and $p_T(\tau_{vis1})/p_T(\tau_{vis2})$
  - e.g. $E_{T,miss}$ ***φ centrality*** between $\tau_{vis1,2}$
  - **for VBF**: exploit topology of forward jets
    - $\Delta\eta(j_1,j_2)$, product $\eta_{j1} x \eta_{j2}$ and inv. mass $m(j_1,j_2)$
    - ***centrality of*** $\tau_{vis1,2}$ ***between leading jets***
- ***In total, 7-9 (6-9) variables for each VBF (Boosted) BDT***

SM: Final BDT Score Distributions

- Careful validation of BDT scores in several background control regions
- Signal region: clear excess of data over expected SM background
  - consistently observed in all channels
- Re-order all BDT score bins through channels and categories according to S/B:

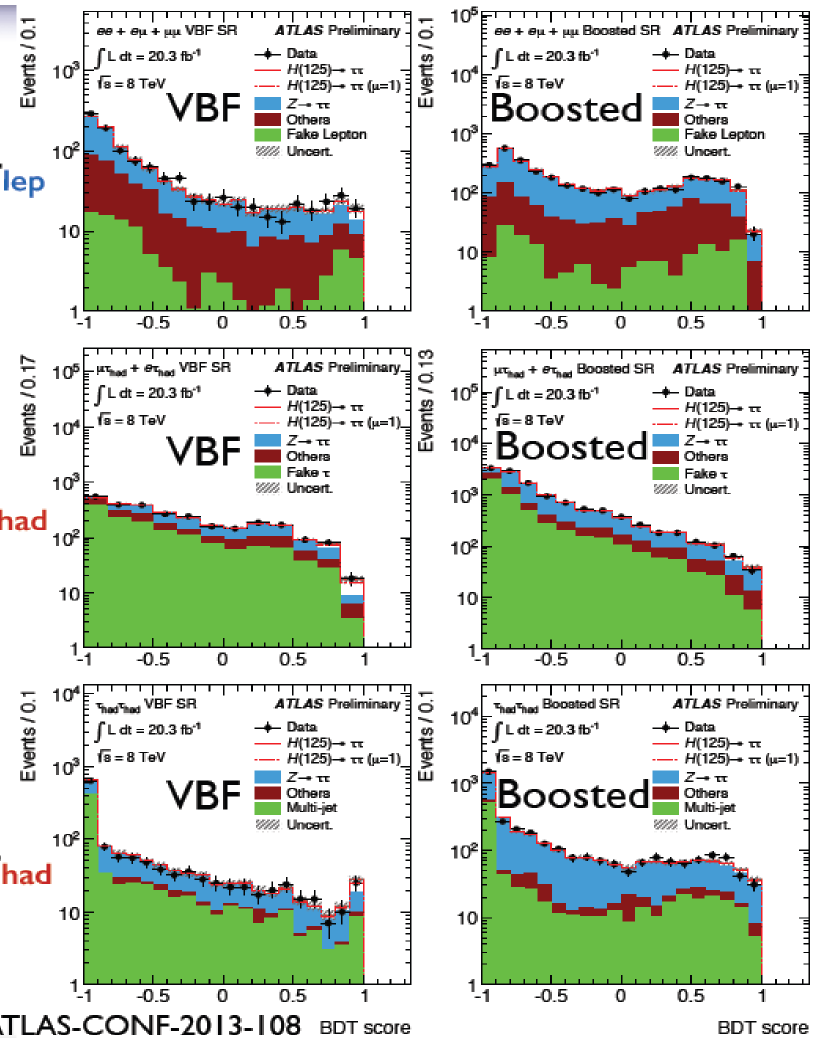


ATLAS-CONF-2013-108

$\tau_{lep}\tau_{lep}$

$\tau_{lep}\tau_{had}$

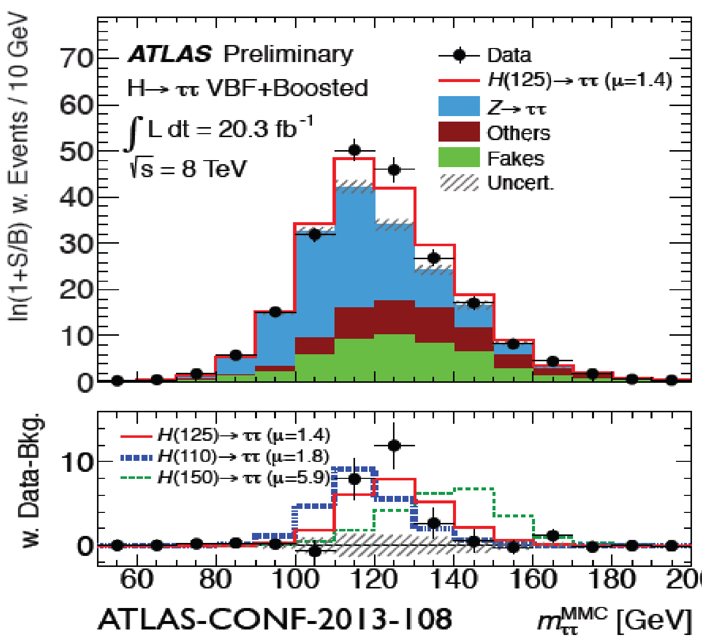
$\tau_{had}\tau_{had}$

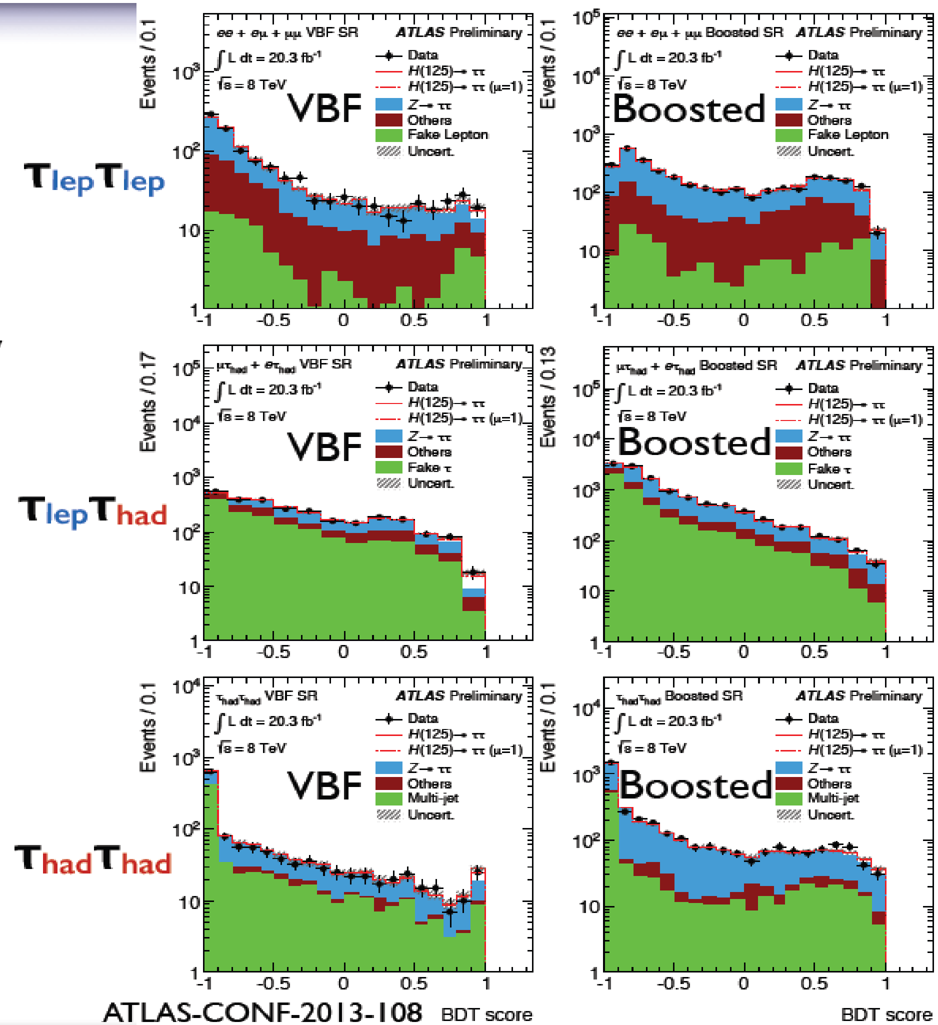


ATLAS-CONF-2013-108

SM: Final BDT Score Distributions

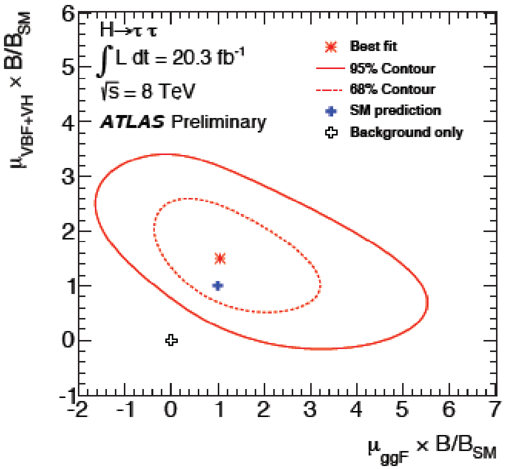
- Careful validation of BDT scores in several background control regions
- Signal region: clear excess of data over expected SM background
  - consistently observed in all channels
- Plot MMC mass weighted with ln(1+S/B) per BDT score bin: consistent with m_H=125 GeV





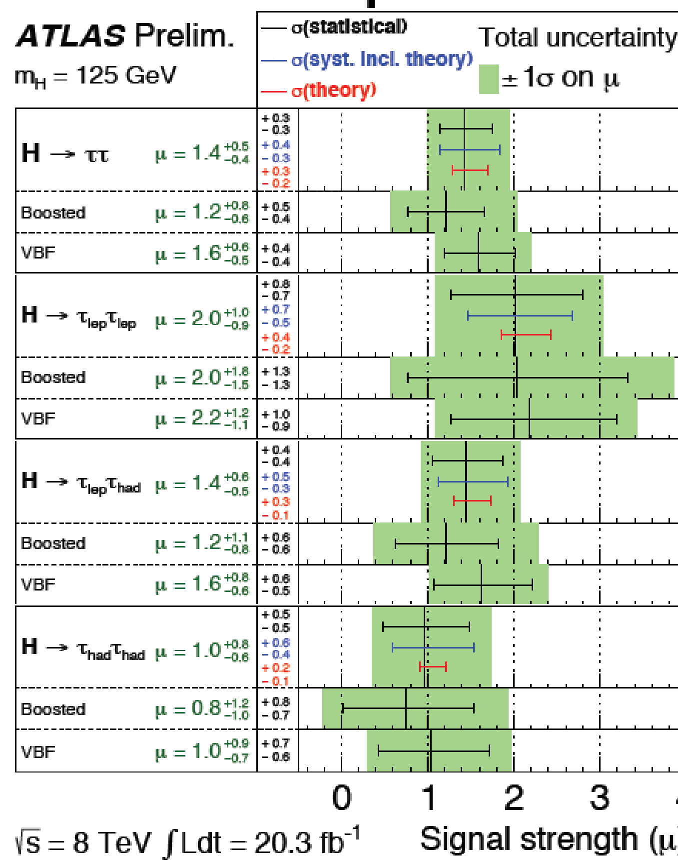
Significance of the Excess

ATLAS-CONF-2013-108



Source of Uncertainty	Uncertainty on μ
Signal region statistics (data)	0.30
$Z \to \ell\ell$ normalization ($\tau_{lep}\tau_{had}$ boosted)	0.13
ggF $d\sigma/dp_T^H$	0.12
JES η calibration	0.12
Top normalization ($\tau_{lep}\tau_{had}$ VBF)	0.12
Top normalization ($\tau_{lep}\tau_{had}$ boosted)	0.12
$Z \to \ell\ell$ normalization ($\tau_{lep}\tau_{had}$ VBF)	0.12
QCD scale	0.07
di-τ_{had} trigger efficiency	0.07
Fake backgrounds ($\tau_{lep}\tau_{lep}$)	0.07
τ_{had} identification efficiency	0.06
$Z \to \tau^+\tau^-$ normalization ($\tau_{lep}\tau_{had}$)	0.06
τ_{had} energy scale	0.06

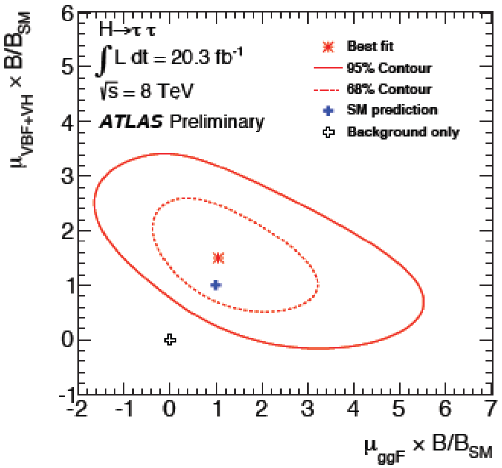
results of separate fits:



- Fit signal in all six BDT score distributions, extract sensitivity (profile likelihood):
  - combined signal strength: $\mu=\sigma/\sigma_{SM}=1.4^{+0.5}_{-0.4}$ at m_H=125 GeV
  - **observed** (expected) **4.1σ** (3.2σ) at m_H=125 GeV
  - consistent with s+b (small „excess“ of 0.9σ dominated by VBF)

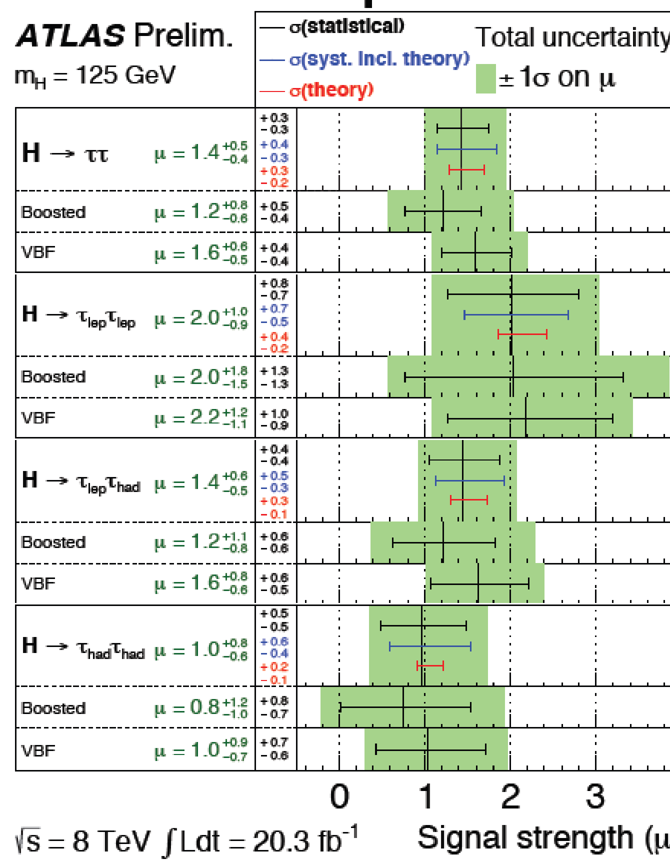
Significance of the Excess

ATLAS-CONF-2013-108



Source of Uncertainty	Uncertainty on μ
Signal region statistics (data)	0.30
$Z \to \ell\ell$ normalization ($\tau_{lep}\tau_{had}$ boosted)	0.13
ggF $d\sigma/dp_T^H$	0.12
JES η calibration	0.12
Top normalization ($\tau_{lep}\tau_{had}$ VBF)	0.12
Top normalization ($\tau_{lep}\tau_{had}$ boosted)	0.12
$Z \to \ell\ell$ normalization ($\tau_{lep}\tau_{had}$ VBF)	0.12
QCD scale	0.07
di-τ_{had} trigger efficiency	0.07
Fake backgrounds ($\tau_{lep}\tau_{lep}$)	0.07
τ_{had} identification efficiency	0.06
$Z \to \tau^+\tau^-$ normalization ($\tau_{lep}\tau_{had}$)	0.06
τ_{had} energy scale	0.06

results of separate fits:



- Fit signal in all six BDT score distributions, extract sensitivity (profile likelihood):
  - combined signal strength: $\mu=\sigma/\sigma_{SM}=1.4^{+0.5}_{-0.4}$ at m_H=125 GeV
  - **observed** (expected) **4.1σ** (3.2σ) at m_H=125 GeV
  - consistent with s+b (small „excess“ of 0.9σ dominated by VBF)
- Evidence for H→ττ (m_H=125 GeV) expected and observed

Summary

- Search for SM $H\to\tau\tau$: BDT-based analysis
  - VBF and Boosted categories for $\tau_{lep}\tau_{lep}$, $\tau_{lep}\tau_{had}$, $\tau_{had}\tau_{had}$
  - observe (expect) 4.1σ (3.2σ) excess over SM background
  - $\mu=\sigma/\sigma_{SM}=1.4^{+0.5}_{-0.4}$ consistent with m_H=125 GeV (μ=1)

ATLAS-CONF-2013-108